

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-188768  
(43)Date of publication of application : 10.07.2001

(51)Int.Cl. G06F 15/18

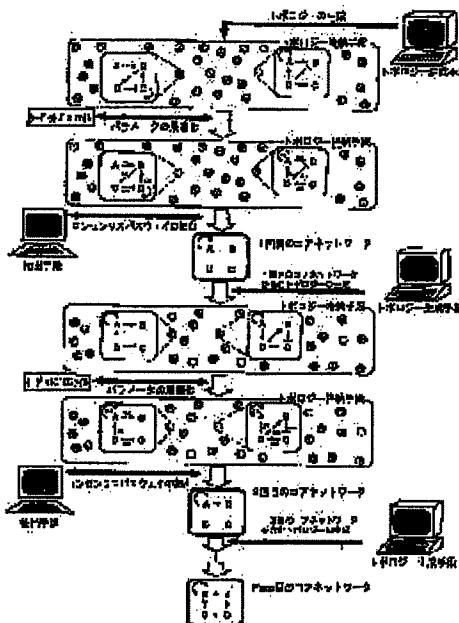
(21)Application number : 11-372936 (71)Applicant : JAPAN SCIENCE & TECHNOLOGY CORP  
(22)Date of filing : 28.12.1999 (72)Inventor : MOROHASHI MINEO  
KITANO HIROAKI

**(54) METHOD FOR ESTIMATING NETWORK**

**(57)Abstract:**

**PROBLEM TO BE SOLVED:** To provide a method for estimating a network based on a data profile.

**SOLUTION:** In this method, a candidate network housing the possibility of reproducing a given data profile is prepared, and the candidate network is evaluated, then the structure and parameter of a network having the possibility of giving the data profile are estimated, and a topology is estimated by stages. Thus, it is possible to estimate the structure of the control network of genes calling the phenomenon or a metabolic network showing the reaction of enzyme or protein from the discovery data of genes or the concentration data of the protein or the like.



(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2001-188768

(P2001-188768A)

(43)公開日 平成13年7月10日 (2001.7.10)

(51)Int.Cl.<sup>7</sup>  
G 0 6 F 15/18

識別記号  
5 5 0

F I  
G 0 6 F 15/18

デーモー<sup>7</sup> (参考)  
5 5 0 C

審査請求 未請求 請求項の数13 O.L (全 15 頁)

(21)出願番号 特願平11-372936

(22)出願日 平成11年12月28日 (1999.12.28)

(71)出願人 396020800

科学技術振興事業団

埼玉県川口市本町4丁目1番8号

(72)発明者 諸橋 峰雄

東京都北区西ヶ原1-27-3-1010

(72)発明者 北野 宏明

埼玉県川越市西小仙波町2-18-3

(74)代理人 100102978

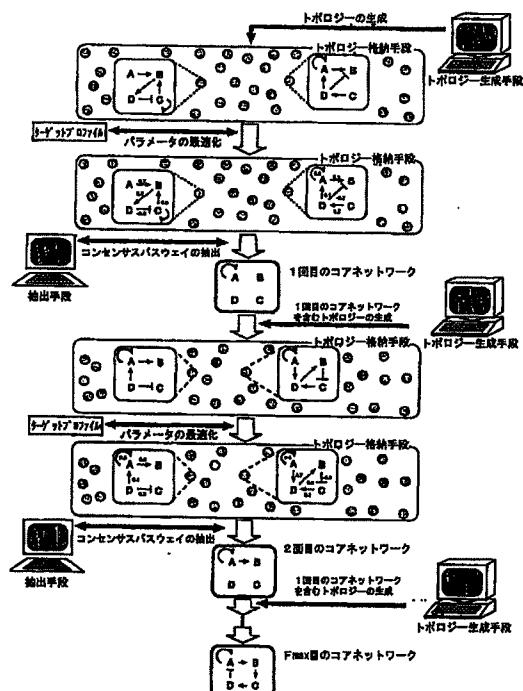
弁理士 清水 初志 (外1名)

(54)【発明の名称】 ネットワーク推定方法

(57)【要約】

【課題】 本発明は、データプロファイルに基づいてネットワークを推定する方法を提供する。

【解決手段】 与えられたデータプロファイルを再現する可能性のある候補ネットワークを作成し、この候補ネットワークを評価し、データプロファイルを与える可能性のあるネットワークの構造と、パラメータを推定し、段階的にトポロジーを推定する方法が提供される。本発明により、遺伝子の発現データやタンパク質の濃度データ等から、その現象を引き起こしている遺伝子の制御ネットワークや酵素やタンパク質の反応を示す代謝ネットワークの構造等を推定することができる。



## 【特許請求の範囲】

【請求項1】 互いに関係している要素間の関係をネットワークとして記述できる場合において、前記要素により生成されたデータが与えられたときに、与えられたデータからそのデータを再現できるネットワークを推定する方法であって、

a) 与えられたデータを再現する可能性の有るトポロジーを生成し、生成された各トポロジーについて、与えられたデータに対する適応度が最も高いパラメータセットを作成して、候補ネットワークを得る工程、

b) 工程a) で作製されたネットワークのうち、目的とする基準を満たすネットワークからコンセンサス・パスウェイを抽出する工程、

c) 工程b) において抽出されたコンセンサス・パスウェイを含むトポロジーを生成し、生成された各トポロジーに対して、与えられたデータに対する適応度が最も高いパラメータセットを与えてネットワークを得る工程、および

d) 工程c) で得たネットワークの規模が目的とする規模に収束するまで、工程c) で得たネットワークを候補ネットワークとして再度工程b) - c) を繰り返して、新たなネットワークを得る工程、とを含むネットワーク推定方法。

【請求項2】 工程b) において、目的とする基準が、適応度によって評価される請求項1に記載のネットワーク推定方法。

【請求項3】 工程b) において、目的とする基準が、そのパスウェイを含むネットワークによって前記生成されたデータに類似し、かつ互いに類似するデータを生成できることである請求項1に記載のネットワーク推定方法。

【請求項4】 工程a) において、与えられたデータを再現可能な全てのトポロジーを生成する工程を含むことを特徴とする請求項1に記載のネットワーク推定方法。

【請求項5】 工程a) において、与えられたデータを再現する可能性の有るネットワークとして、前記可能性の高いネットワークを選択する工程を含むことを特徴とする請求項1に記載のネットワーク推定方法。

【請求項6】 与えられたデータとの誤差が少ないデータを再現するネットワークを生成する工程を含むことを特徴とする請求項5に記載のネットワーク推定方法。

【請求項7】 工程a) および/または工程c) におけるネットワークを得る工程が、与えられたデータとの誤差が少ないデータを再現するネットワーク間でその構造の一部を組み換えて、ネットワークを作成する工程を含むことを特徴とする請求項5または6に記載のネットワーク推定方法。

【請求項8】 工程a) および/または工程c) におけるネットワークを得る工程が、任意ネットワークの構造的近傍を探して、ネットワークを作成する工程を含むこ

とを特徴とする、請求項5～7のいずれかに記載のネットワーク推定方法。

【請求項9】 工程a) および/または工程c) におけるネットワークを得る工程が、同一のネットワークに対して1個のパラメータ、または複数であって互いに異なるパラメータのセットを推定することを特徴とする、請求項1～8のいずれかに記載のネットワーク推定方法。

【請求項10】 工程a) および/または工程c) における候補ネットワークを得る工程が、遺伝的アルゴリズム、シミュレーテッド・アニーリング、および山登り法からなる群から選択されるすくなくとも一つの手法を用いてパラメータを推定する工程を含むことを特徴とする、請求項1～8のいずれかに記載のネットワーク推定方法。

【請求項11】 前記要素によって生成されたデータが、遺伝子の発現プロファイルである請求項1～10のいずれかに記載のネットワーク推定方法。

【請求項12】 互いに関係している要素間の関係をネットワークとして記述できる場合において、前記要素により生成されたデータが与えられたときに、与えられたデータからそのデータを再現できるネットワークを推定するネットワーク推定装置であって、

a) トポロジーと、それに対応するパラメータ・セットとで構成されるネットワークを記憶する第1の記憶手段と、

b) 選択された候補としてのトポロジー、またはその要素と、それに対応するパラメータ・セットを格納する第2の記憶手段と、

c) 与えられたデータを再現する可能性のあるトポロジーを生成し、生成された各トポロジーについて、与えられたデータに対する適応度が最も高いパラメータ・セットを作成し、複数の候補ネットワークを得て、前記第1の記憶手段に格納する手段と、

d) 第1の記憶手段に格納された候補ネットワークのうち目的とする基準を満たすネットワークからコンセンサス・パスウェイを抽出して、前記第2の記憶手段に格納する手段と、

e) 手段d) によって第2の記憶手段に格納されたコンセンサス・パスウェイを含むトポロジーを生成し、生成された各トポロジーについて、与えられたデータに対する適応度が最も高いパラメータセットを与えてネットワークを得て、前記第1の記憶手段に格納する手段、および

f) 推定すべきネットワークが得られるまで、手段e)において第1の記憶手段に格納されたネットワークを候補ネットワークとして手段d) における候補ネットワークとして与える手段、とを含む装置。

【請求項13】 請求項1～11のいずれかに記載のネットワーク推定方法をコンピュータ・システムに実現さ

せることができるプログラムを格納した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、互いに関連し、干渉し合う別々の要素のネットワークとしてモデル化できるシステムを、そのネットワークの挙動や状態を表す情報から逆に推定する方法に関する。本発明のネットワーク推定方法は、生物学的現象における遺伝子の発現データ、タンパク質の濃度データ等から、その現象を引き起こしている遺伝子の制御ネットワークや酵素やタンパク質の反応を示す代謝ネットワークの構造等の推定に有用である。

【0002】

【従来の技術】最近の分子生物学の急激な進歩によって、様々な生物に関する多くの情報が蓄積されつつある。このような生物の典型例が、酵母、線虫、あるいはショウジョウバエ等である。例えば、線虫 (*C.elegans*) では、既にその全細胞系譜が推定され (Salton et al. Dev.Biol.100,64-119,1983)、神経系の結合などの同定も行われている (White et al. Phil.Trans.R.Soc.314, 1-340, 1986)。また1998年には、ゲノムDNAの全塩基配列が明らかにされた (Science 282, 2012-2017, 1998)。線虫の他、酵母、大腸菌、マイコプラズマ等のモデル生物においても、ゲノムDNAの全塩基配列がすでに決定されている。そして究極の目標であるヒト・ゲノムDNAの全塩基配列も、数年以内には終了すると言われている。しかしながら、塩基配列が明らかにされること、遺伝子やその機能、さらには遺伝子間の相互作用を明らかにすることは、まったく性格の異なる課題である。すなわち、ゲノムの塩基配列の決定は、単に遺伝子の物質としての側面を明らかにしたということにすぎない。生物を遺伝情報に基づいて理解するには、決定された塩基配列に含まれる個々の遺伝子の働きと相互の関連を明らかにする必要がある。

【0003】遺伝子の機能や遺伝子間の相互作用を明らかにすることは、非常に困難な仕事である。たとえば平均的なヒトの細胞を構成するタンパク質の種類は、5000～7000とも言われている。これほど多様なタンパク質の一つ一つを同定し、その機能を明らかにし、更に相互関係を見出して、初めて集合体 (プロテオーム; proteome) として理解することができる。現在行われている分子生物学の研究の多くは、注目する現象に関わる遺伝子およびその転写産物の特定と、その現象を引き起こす相互作用の解明に重点を置いている。しかし遺伝子間の相互作用を理解するには、多くの遺伝子間を結びつける膨大な組み合わせを考慮しなければならない。このような作業を人間の能力で行うことは極めて難しい。我々は、この困難さをコンピュータという計算能力をもった道具を導入することで乗り越えようとしている。

【0004】この考えを図式したのが図1である。これ

は、コンピュータを利用する部分 (パーキャル) と、実際に生物学的に実験する部分 (リアル) との関係を示す図である。図1において、まず、解明しようとする現象を特定すると、現在知られている生物学的知識 (図1 フェーズ-I A) から、モデルをコンピュータ上に実装する (図1 フェーズ-I B)。また、すでに仮説が提示されている場合には、その仮説を実装する。実装されたモデルを使ってシミュレーションを行い (図1 フェーズ-I C)、観測されているデータとの整合性を検証する (図1 フェーズ-I D)。シミュレーションの結果と、観測データと一致しない場合は、2つの可能性が考えられる。まず、シミュレーションが正確でない場合である。これは、シミュレーションの精度をあげて、確実なものとすることで解決する。次に考えられるのは、モデルが不完全または間違っている場合である。この場合は知られている実験データに整合するモデルを構築する必要がある。

【0005】その後に、できるならば、このモデルを用いて、コンピュータ上で実験を行う (図1 フェーズ-I I F)。フェーズI I の後、又はフェーズI I ができないなら、直接、仮説的に決めたモデル等が正しいか実際に生物学的な実験を開始する (図1 フェーズ-I I I およびI V)。このようにして、「生物のリバース・エンジニアリング」を行うことができる。コンピュータ上におけるシミュレーションにより、仮説的に求めたモデルの絞り込みを必要としている分野は、生物学分野に限らない。

【0006】例えば分子生物学においては、DNAチップやマイクロアレイ、さらには、PCRなどをを利用して得られた遺伝子産物やタンパク質の量の時間的変化に関するデータに加え、電気生理学的実験や電圧感受性色素などを用いた神経回路網の活動に関するデータ等に基づくネットワークの推定が行われる。これらのデータの背後にある遺伝子やタンパク質、酵素の相互作用のネットワーク、および、神経細胞のネットワークで表されるモデルを、コンピュータを用いて推定することができれば有用である。しかし現実には、与えられたデータから考えられるモデルの総数は莫大な数である。従来は、人間の限られた処理能力で評価できる範囲のモデルのみが考慮されていた。あるいは半ば人間の勘に頼ってあたりを付け、生物学的な実験により確かめていた。これに対して本発明者らが目指しているのは、与えられたデータから、類似したデータを生成する可能性のあるモデル (ネットワーク) を、コンピュータを用いて推定することにより、確からしいモデルの数をある程度まで絞り込むという戦略である。これは、上述の図1に示した場合は、フェーズI に関するものである。

【0007】本発明において確からしいモデルとは、ネットワークを推定するときに手にしているデータに基づいて推測することができる合理的なモデルを意味する。

合理的なモデルは、次の実験計画を考えたり、あるいは新たな仮説を構築する上で有用である。

【0008】これらの相互作用のモデルの推定（ネットワークの推定）は、疾病の原因遺伝子の特定や、投薬の効果の推定等に必要である。また、未知の遺伝子や遺伝子産物の推定にも用いることができる。更に生物学的な分野での応用に限らず、たとえば、交通ネットワークや経済システム等の分野においても、コンピュータを利用したシミュレーションに基づく仮説モデルの検証は、重要な解析手法となっている。生物学的な現象のみならず、コンピューターネットワークのような、人工的なネットワークも解析の必要なネットワークとして挙げることができる。たとえば、ネットワークのネットワークと呼ばれるインターネットにおいては、トライフィック（ネットワークを流れた情報量）に関するシミュレーションがしばしば行われる。ルータ等に対するアクセス頻度はもとより、流れるパケットの量やサイズなど、多くのデータからスループットの予測などが行われている。これらの解析行為は、ネットワークの推定に他ならない。

【0009】本発明者らも、このような考え方に基づいてネットワークの推定方法を確立している。すなわち、ターゲット・プロファイルを再現することができる候補ネットワークを生成し、突然変異アナリシスによってそれをスクリーニングする方法を提案した(Morohashi et al. Proc. of European Conference on Artificial Life, 1999)。ここで、ターゲット・プロファイルとは、明らかにすべきネットワークによって現実にもたらされたデータの集合である。また突然変異アナリシスは、変異を含むネットワークによって与えられたデータに基づいて、候補ネットワークの評価を行う手法である。この方法においては、ターゲット・プロファイルや突然変異アナリシスのためのデータは、現実の実験によって集めなければならない。つまり、「リアル」である。したがって、突然変異アナリシスのための実験を繰り返し行う必要があった。またこの方法では、単一の候補をスクリーニングする工程を伴っていた。そのため、選択の対象となる範囲を広く取りにくいという問題があった。

【0010】

【発明が解決しようとする課題】本発明は、コンピュータにより実施することができ、仮説的に求めたモデルの絞り込みを効率良く行うことが可能なネットワーク推定方法の提供を課題とする。より具体的には、現実の実験を少なくしながら、しかも効率的にネットワークの推定を行うことができる方法の提供が本発明の課題である。

【0011】

【課題を解決するための手段】本発明者らは、先に提案したネットワークの推定方法を構成する各ステップの中で、突然変異アナリシスの存在に着目した。コンピューターによるシミュレーションを有効に活用するには、現実に実験によって得なければならないデータを最小限に

することが重要である。言いかえれば、少ない実験データに基づいて、できるだけ多くのモデルをスクリーニングし、そして現実との誤差の小さいネットワークを選び出すことができれば理想的である。このような考え方に基づいて、突然変異アナリシスのような実験を伴う評価手法を必要としない新たなネットワーク推定方法を実現するための研究を重ねた。

【0012】本発明者らは、候補ネットワークで構成されるライブラリーの中から、特にターゲット・プロファイルとの誤差が小さいネットワークに着目した。そして、それらが共通のパスウェイを高い頻度で含むことを見出した。なおトポロジーとは、ネットワークを構成する物理的な構造を意味する。トポロジーはパスウェイの集合によって構成される。パスウェイとは、ネットワークを構成する要素を結びつける線に相当する。したがってネットワークとは、パスウェイの集合体であるトポロジーに、様々なパラメータを伴ったものと言うことができる。更に本発明者らは、この共通のパスウェイを抽出し、得られたパスウェイをトポロジーに含む新たな候補ネットワークを生成して再びそれを評価することを繰り返すことによって、最終的にデータを再現することができるネットワークの構成が可能となることを見出して本発明を完成した。すなわち本発明は、以下のネットワーク推定方法と、そのための装置に関する。

〔1〕互いに関係している要素間の関係をネットワークとして記述できる場合において、前記要素により生成されたデータが与えられたときに、与えられたデータからそのデータを再現できるネットワークを推定する方法であって、

a) 与えられたデータを再現する可能性の有るトポロジーを生成し、生成された各トポロジーについて、与えられたデータに対する適応度が最も高いパラメータセットを作成して、候補ネットワークを得る工程、

b) 工程a) で作製されたネットワークのうち、目的とする基準を満たすネットワークからコンセンサス・パスウェイを抽出する工程、

c) 工程b) において抽出されたコンセンサス・パスウェイを含むトポロジーを生成し、生成された各トポロジーに対して、与えられたデータに対する適応度が最も高いパラメータセットを与えてネットワークを得る工程、および

d) 工程c) で得たネットワークの規模が目的とする規模に収束するまで、工程c) で得たネットワークを候補ネットワークとして再度工程b) - c) を繰り返して、新たなネットワークを得る工程、とを含むネットワーク推定方法。

〔2〕工程b) において、目的とする基準が、適応度によって評価される〔1〕に記載のネットワーク推定方法。

〔3〕工程b) において、目的とする基準が、そのパス

ウェイを含むネットワークによって前記生成されたデータに類似し、かつ互いに類似するデータを生成できることである〔1〕に記載のネットワーク推定方法。

〔4〕工程a)において、与えられたデータを再現可能な全てのトポロジーを生成する工程を含むことを特徴とする〔1〕に記載のネットワーク推定方法。

〔5〕工程a)において、与えられたデータを再現する可能性の有るネットワークとして、前記可能性の高いネットワークを選択する工程を含むことを特徴とする〔1〕に記載のネットワーク推定方法。

〔6〕与えられたデータとの誤差が少ないデータを再現するネットワークを生成する工程を含むことを特徴とする〔5〕に記載のネットワーク推定方法。

〔7〕工程a)および/または工程c)におけるネットワークを得る工程が、与えられたデータとの誤差が少ないデータを再現するネットワーク間でその構造の一部を組み換えて、ネットワークを作成する工程を含むことを特徴とする〔5〕または〔6〕に記載のネットワーク推定方法。

〔8〕工程a)および/または工程c)におけるネットワークを得る工程が、任意ネットワークの構造的近傍を探して、ネットワークを作成する工程を含むことを特徴とする、〔5〕～〔7〕のいずれかに記載のネットワーク推定方法。

〔9〕工程a)および/または工程c)におけるネットワークを得る工程が、同一のネットワークに対して1個のパラメータ、または複数であって互いに異なるパラメータのセットを推定することを特徴とする、〔1〕～〔8〕のいずれかに記載のネットワーク推定方法。

〔10〕工程a)および/または工程c)における候補ネットワークを得る工程が、遺伝的アルゴリズム、シミュレーテッド・アニーリング、および山登り法からなる群から選択されるすくなくとも一つの手法を用いてパラメータを推定する工程を含むことを特徴とする、〔1〕～〔8〕のいずれかに記載のネットワーク推定方法。

〔11〕前記要素によって生成されたデータが、遺伝子の発現プロファイルである〔1〕～〔10〕のいずれかに記載のネットワーク推定方法。

〔12〕互いに関係している要素間の関係をネットワークとして記述できる場合において、前記要素により生成されたデータが与えられたときに、与えられたデータからそのデータを再現できるネットワークを推定するネットワーク推定装置であって、

a) トポロジーと、それに対応するパラメータ・セットとで構成されるネットワークを記憶する第1の記憶手段と、

b) 選択された候補としてのトポロジー、またはその要素と、それに対応するパラメータ・セットを格納する第2の記憶手段と、

c) 与えられたデータを再現する可能性のあるトポロジ

10

20

30

40

50

ーを生成し、生成された各トポロジーについて、与えられたデータに対する適応度が最も高いパラメータ・セットを作成し、複数の候補ネットワークを得て、前記第1の記憶手段に格納する手段と、

d) 第1の記憶手段に格納された候補ネットワークのうち目的とする基準を満たすネットワークからコンセンサス・パスウェイを抽出して、前記第2の記憶手段に格納する手段と、

e) 手段d)によって第2の記憶手段に格納されたコンセンサス・パスウェイを含むトポロジーを生成し、生成された各トポロジーについて、与えられたデータに対する適応度が最も高いパラメーターセットを与えてネットワークを得て、前記第1の記憶手段に格納する手段、および

f) 推定すべきネットワークが得られるまで、手段e)において第1の記憶手段に格納されたネットワークを候補ネットワークとして手段d)における候補ネットワークとして与える手段、とを含む装置。

〔13〕〔1〕～〔11〕のいずれかに記載のネットワーク推定方法をコンピュータ・システムに実現させることができるプログラムを格納した記録媒体。

### 【0013】

【発明の実施の形態】本発明におけるネットワークは、トポロジー、パラメータ、および適応度の3つの成分で構成される。これら3つの成分をトリプレットと呼ぶ。要素はネットワークを構成している物質を意味し、その関係を表すのがトポロジーである。トポロジーは要素間のつながりを示す物理的な構造からなり、この構造はパスウェイに分解することができる。逆にトポロジーは、パスウェイの集まりによって表現することができる。パスウェイの中でも、特に複数のトポロジーに共通して見出されるパスウェイを、コンセンサス・パスウェイと呼ぶ。本発明におけるトポロジーは、ネットワークによって現実のデータを再現することができるようモデリングされる。モデリングに必要な全ての数値はパラメータと呼ばれる。モデリングとは、物理的な構造のみから構成されるトポロジーに量的な概念を与える操作に他ならない。一方、現実にネットワークによって生成されたデータの集合は、ターゲット・プロファイルと呼ぶ。パラメータを伴ったトポロジーは量的な概念を伴っているので、ターゲット・プロファイルとの誤差を算出することができる。誤差の大小はターゲット・プロファイルとの類似度を意味し、本発明ではこれを適応度と呼ぶ。

【0014】本発明によって提供されるネットワーク推定方法は、以下の工程を含む。なお本発明においては、互いに関係している要素間の関係がネットワークとして記述できることが条件である。したがってターゲット・プロファイルは、要素間の相互作用の結果として生成されたものであると言うことができる。本発明のネットワークの推定方法は次の工程a)～d)を含むことを特徴

とし、ターゲット・プロファイルを再現しうるネットワークの推定を目的としている。

a) 与えられたデータを再現する可能性の有るトポロジーを生成し、生成された各トポロジーについて、与えられたデータに対する適応度が最も高いパラメータセットを作成して、候補ネットワークを得る工程、

b) 工程a) で作製されたネットワークのうち目的とする基準を満たすネットワークからコンセンサス・パスウェイを抽出する工程、

c) 工程b) において抽出されたコンセンサス・パスウェイを含むトポロジーを生成し、生成された各トポロジーについて、与えられたデータに対する適応度が最も高いパラメータセットを与えてネットワークを得る工程、および

d) 工程c) で得たネットワークの規模が目的とする規模に収束するまで、工程c) で得たネットワークを候補ネットワークとして再度工程b) - c) を繰り返して、新たなネットワークを得る工程、

【0015】本発明において、互いに関係している要素間の関係がネットワークとして記述できることとは、ネットワークを構成する要素の少なくとも一部は、相互に関連性を持っていることを意味する。このようなネットワークとしては、たとえばある細胞を構成する遺伝子間の相互作用を示すことができる。この場合、細胞に含まれる遺伝子の発現プロファイルが本発明のターゲット・プロファイルに相当する。遺伝子の発現プロファイルは、遺伝子の発現レベルを、mRNAやタンパク質の濃度として解析することによって得ることができる。より具体的には、たとえばDNAチップを用いて、mRNAの発現プロファイルを得ることができる。DNAチップを利用することにより、一定のノード（つまりDNAチップに固定されたプローブセット）に対する発現レベルを、効率的に収集することができる。あるいは、二次元電気泳動によって、タンパク質の発現プロファイルを得る方法も公知である。ターゲット・プロファイルを再現するべきネットワークは、遺伝子をノードとするトポロジーに相当する遺伝子間の関連性、相互作用の程度を表すパラメータ、そして発現プロファイルとの類似性を表す適応度で構成されるトリプレットによって表すことができる。図4にトポロジーを図示した。図4a)において、要素A B C Dは、A→B→C→Dの活性化の関係、並びにD-| B間の抑制関係という相互の関連性、そしてAの自己活性化という関係にあることを示している。本発明によって推定することができるネットワークと、それを構成するトリプレットを以下に例示する。

【0016】たとえば酵素の相互作用のネットワークは、基質をノードとする酵素作用がトポロジーとなり、その活性の強さや阻害の程度がパラメータに還元される。この場合の適応度は、細胞や組織における代謝マップとの類似性となる。電気生理学的実験や電圧感受性色

素などを用いた神経回路網の活動に関するデータでは、神経細胞間のシナプス結合がトポロジーに相当し、そこに流れる伝達シグナルがパラメータを構成する。この場合の適応度は、神経回路網の活動状態との類似性を意味する。更に、交通ネットワークとそれを結びつけるターミナル、あるいは経済システムとそれを構成する要素なども、本発明におけるネットワークとして記述することができる。

【0017】通常、ネットワークを構成する要素のうちのどれだけのものが関連性を持っているのかは未知である。また要素間の関連性は、完全に未知であっても良いし、あるいは部分的に確認されたものであってもよい。

【0018】工程a) において生成するトポロジーは、一般的には、できるだけ広い範囲から絞り込みが行えるように、多様な組み合わせを生成するのが望ましい。例えば、想定することができるトポロジーの全てを生成することは、本発明の好ましい態様の一つである。しかしながら、考慮すべき要素が多い場合には、要素の組み合わせの全てをカバーできる想定可能なトポロジーを生成することは困難である。トポロジーの候補があまりにも巨大になる場合には、スクリーニングを困難とする場合がある。したがって、必要に応じてトポロジーを選択することもできる。トポロジーの選択は、想定可能な全てのトポロジーから例えばランダムに選択することによって行うことができる。工程a) で生成されるトポロジーの集合をトポロジープールと呼ぶ。

【0019】また、ネットワークを構成する要素間の関連性が予め部分的に明らかな場合には、これを固定した上でトポロジーを生成することによって、合理的なスクリーニングを行うことができる。たとえば細胞の遺伝子発現プロファイルの解析においては、通常、いくつかの遺伝子の間の相互関係が部分的に明らかにされている場合が多い。このような場合には、既に明らかな部分については、固定した状態でトポロジーを生成させるのが合理的である。あるいは逆に、あえて通説となっている要素間の関係を無視してトポロジーを生成することにより、既知の関係とは全く異なった相互作用の可能性を見出すことが期待できる。更に、ネットワークを構成する要素の相互の関連性が未知であっても、その一部の関連性を仮定に基づいて実験的に固定したトポロジーを生成することもできる。このようにして生成した様々なトポロジーに本発明を適用することによって、その仮定の確からしさを相互に比較することができる。

【0020】このとき、本発明によるネットワークの推定方法では、複数の候補ネットワークを生成することができることから、考え得る候補ネットワークを幅広く把握することができる。このことは、単一の候補ネットワークの推定を目指す公知の方法に対して大きなメリットである。また、様々な条件の基で推定した候補ネットワークの比較を、より有効に行うことができる。たとえ

ば、ある条件の基で推定したときに得られた候補ネットワークのランキングと、他の条件における候補ネットワークのランキングを比較することができる。

【0021】工程a)におけるトポロジーやネットワークの選択は、スクリーニングすべき対象を狭める恐れがあるのは事実である。しかしながら次のような理由により、工程a)における候補の選択が、ネットワークの推定において最終的な結果に対して大きな影響を与えることはない。本発明においては、後に述べるように、データを再現する可能性のあるトポロジーを段階的に推定する工程が繰り返される。したがって、たとえばある段階で、データの再現が可能な特定の候補ネットワークが候補集団から失われたとしても、その影響は部分的な影響にとどまる。なぜならば、本発明においては、高頻度に見られるコンセンサス・パスウェイの抽出工程が繰り返される。したがって、候補ネットワークがランダムに選択される場合には、特定のトポロジーの有無に関わらず、コンセンサス・パスウェイの抽出は可能なのである。

【0022】更に本発明では、コンセンサス・パスウェイを含むトポロジーを生成し、生成されたトポロジーにパラメータを与えて再び候補ネットワークを得る工程を含む。このとき生成される候補ネットワークは、コンセンサス・パスウェイを含むことを条件として、無作為に想定されるトポロジーの全てについて生成することもできるし、あるいは前記与えられたデータとの誤差が少ないデータを再現するネットワークを高頻度で作成するステップを含むこともできる。

【0023】目的とするネットワークを高頻度で作成するステップにより、与えられたデータを再現する可能性の高い候補ネットワークを選択する確率が高くなる。また、前記与えられたデータとの誤差が少ないデータを再現するネットワーク同士の一部を組み替えて、新たなネットワークを作成するステップを含むことができる。これは遺伝的アルゴリズム(GA)等を用いることにより実現することができる。

【0024】遺伝的アルゴリズムとは、問題解決手法の一つとして考え出されたアルゴリズムである。遺伝的アルゴリズムにおいては、無作為に生成した候補集団を、特定の視点で評価し、一定の水準にある候補のみを選択し、選択された候補に基づいて新たな候補集団を生成する工程を繰り返すことによって、ある目的を達成することができる候補が作り出される。その工程が、あたかも生物の進化の原理を見るようであることから、遺伝的アルゴリズムと呼ばれている。本発明に遺伝的アルゴリズムを適用すれば、たとえば適応度が高い候補ネットワークの間で、人為的な組み換えを生成することができる。組み換えによって適応度に優れたネットワークが多く得られる場合には、同様の組み換えが遺伝的アルゴリズムに基づいて更に積極的に行われるであろう。

【0025】こうして得ることができる候補ネットワークは、ネットワーク空間を広く粗く探索して得られている。ここでネットワーク空間とは、想定されるネットワークの全てが含まれる空間を意味している。ネットワーク空間の広く粗い探索は、広域探索と呼ばれる。広域探索を補うために、ネットワークを得るための探索方法として、任意のトポロジーの構造的近傍を探索して、トポロジーを作成するステップを含むことができる。このような検索方法は、広域探索に対して局所探索と呼ばれる。局所探索は、例えば、シミュレーテッド・アニーリング(SA)を用いて行うことができる。候補ネットワークの生成に、広域探索および局所探索をハイブリッドで用いることにより、最適なネットワークを効率的に選択することができる。

【0026】工程a)においてネットワークを得るステップは、同一のトポロジーに対して、1個のパラメータ、若しくは、複数であって互いに異なるパラメータのセットを推定することができる。パラメータの推定には、例えば、遺伝的アルゴリズム(GA)、シミュレーテッド・アニーリング(SA)、あるいは山登り法等の公知の最適化手法(「遺伝的アルゴリズム」監修:北野宏明、産業図書発行、平成5年)を用いることができる。これにより、与えられたデータを再現するための、トポロジーに対する最適なパラメータを早く推定することができる。山登り法(Hill Climbing:HC)とは、可能性のある探索点の中で最も有望な点を選んで探索を進めていく方法である。探索点の近傍の中から最適の解を見つける方法である。SAは局所解(local minimum)に陥るのを防ぐために、HCの考え方たに更に確率的な遷移を導入している。局所解とは、限られた範囲では最適な解ではあるが、広い探索範囲においては最適とは言えない解を意味する用語である。HCは、局所的な比較を繰り返すことで解を導くアルゴリズムなので、場合により局所解に陥りやすい傾向がある。さてSAでは、探索点X<sub>i</sub>の近傍を探索して解Xが得られたときに、その評価値がX<sub>i</sub>より悪い場合でも、X<sub>i+1</sub> = Xとなる可能性を残す点においてHCと相違している。つまり、解が改善された場合には解を置きかえ、更に改善されない場合にも一定の確率で置き換える。このような操作によれば、探索点の数よりも多くの解が残されるとともに、局所解のみならず、より最適解に近い解を得られる可能性が高まる。

【0027】工程b)においては、先に生成された候補ネットワークのうち目的とする基準を満たすネットワークからコンセンサス・パスウェイが抽出される。本発明において満たすべき基準とは、推定すべきネットワークによって生成されたデータとの誤差が許容範囲内にあることを意味する。本発明において、推定すべきネットワークによって生成されたデータを、特にターゲットプロファイルと呼ぶ。各ネットワークとターゲットプロファ

イルとの誤差は、任意の方法によって比較することができる。たとえば、ネットワークの適応度を比較することによって、候補ネットワークとデータとの誤差の大きさをランクイングすることができる。通常コンセンサス・パスウェイは、適応度の評価にしたがってランクイングした候補ネットワークの、上位候補に高頻度に見出される構造として抽出される。本発明におけるコンセンサス・パスウェイとは、複数のネットワーク候補から抽出することができる、トポロジーの部分的な構造を意味する。先に述べたように、本発明において、コンセンサス・パスウェイは、単一であることもできるし、複数を選択することもできる。たとえば、あるネットワーク候補の集団において、目的とする基準を満たすネットワークから論理的に相反する複数の構造が抽出された場合には、抽出された複数の構造をそれぞれ候補として採用し、以後の工程を各候補ごとに分割して進めることができる。本発明において、コンセンサス・パスウェイを抽出し、以降の工程でこのコンセンサス・パスウェイに基づいて、更に次のコンセンサス・パスウェイの抽出操作を繰り返すとき、先に抽出したコンセンサス・パスウェイを固定すると表現する。

【0028】上位候補とは、最小二乗法による適応度で、たとえば上位5%の集団が望ましい。このような集団における高頻度とは、たとえば5%以上、より好ましくは3%以上の確率で見出される場合を言う。このとき、予め設定された上位5%という選択基準を、状況によって変動させることもできる。たとえば、コンセンサス・パスウェイとして抽出すべき要素が多くのパターンに分散し、かつ上位5%のネットワークとそれ以下のネットワークの間で適応度に優位な差が見出せない場合には、上位5%のみでは選択範囲として不十分となる恐れがある。このようなケースでは、適応度に有意な差を見出すことができるところまで、コンセンサス・パスウェイの選択範囲を広げると良い。逆に、コンセンサス・パスウェイとして抽出すべき要素が限られたパターンに集中するようなケースでは、不必要に選択範囲を広げる必要は無い。

【0029】さて、本発明において適応度の計算根拠であるターゲット・プロファイルは、単一であることもできるし複数を利用することもできる。たとえば、正常なネットワーク(wild type)において生成されたデータに加え、ネットワークを構成する要素の一部を人為的に破壊したネットワーク(mutant)によってもたらされたデータからなるターゲット・プロファイルを組み合わせることもできる。複数のターゲット・プロファイルを用いることによって、本発明による候補ネットワークの評価は多面的に行われることになり、結果として候補の絞込みを行いやすくなる。具体的には、たとえばmutantに基づくターゲット・プロファイルの参照によって、論理的にあり得ない候補を評価の対象から外すことができる。

るいは、より確からしい候補を高く評価することもできるようになる。

【0030】コンセンサス・パスウェイを選択するためのネットワークが満たすべき基準は、適応度の比較の他に、あるパスウェイを含むネットワークが、ターゲットプロファイルに類似し、かつ互いに類似するデータを生成できるかどうかを評価することによって判断することができる。コンセンサス・パスウェイを含むネットワークによって生成されるデータは、ターゲット・プロファイルに類似するデータとなるべきである。したがって、あるパスウェイを含むネットワークが、ターゲット・プロファイルに類似するデータを生成できるかどうかを予測すれば、そのパスウェイをコンセンサス・パスウェイとして抽出すべきかどうかを判断することができる。そのためには、たとえば次のようなアルゴリズムによってコンセンサス・パスウェイを固定することもできる。この手法の重要な点はいかにコンセンサス・パスウェイを導出するかにある。ここではその基準として $N \times N$ 行列(ネットワーク構成要素数がNの場合)のベクトル $\Gamma$

(ガンマ)を定義する。トポロジープールの中で適応度の高いエリート数Tのトポロジーを用いて、 $\Gamma$ の各要素 $\gamma_{ij}$ を以下に示す式1で求める。

【0031】

【数1】式1

$$\gamma_{ij} = \sum_{k=1}^T w_{ij}^k \times A_i \times \beta_{ij}$$

【0032】ここで、 $w_{ij}^k$ はトポロジーkのi行j列要素の結合荷重値、 $A_i$ はノードiの全ステップを通した平均遺伝子産物濃度、 $\beta_{ij}$ はエリート中のi行j列要素の結合状態値(1:結合、0:非結合)の総和を示している。

【0033】この式1では結合荷重値、ノードの発現産物濃度およびノード間の結合頻度を反映した結果を導き出しが、これだけだと強い影響力を持つ結合のみを求めることになり、ありえる非結合を導くことができない。そこで、さらに非結合に関して基準 $\gamma$ を導入する。その各要素 $\gamma_{ij}$ を以下のように定義する。

【0034】

【数2】式2

$$\gamma'_{ij} = \frac{1}{\sum_{k=1}^T w_{ij}^k + \varepsilon} \times A_i \times (T - \beta_{ij})$$

ここで、 $\varepsilon$ は定数( $\approx 0$ )である。

【0035】基準行列 $\Gamma$ を導出した時点で、この中で最も大きい値を持つ要素がもっともあり得るパスウェイとして考えられる。すなわち、例えば $\Gamma$ (activation)の中で $\gamma(2,3)$ の値が最も大きければノード2からノード3への活性制御が大きく効いていると言える。よって、この $\Gamma$ をもとに「1サイクルで固定するパスウェイ

の数」をあらかじめ決めておき、以下に説明している探索経路などを利用して再帰的にコンセンサス・パスウェイを固定していくことができる。本発明における1サイクルとは、前記工程a)～c)によって候補ネットワークの生成と評価を経てコンセンサス・パスウェイを抽出し、このコンセンサスパスウェイに基づいて新しいネットワークを生成するまでを言う。後に述べるように、本発明によるネットワークの推定方法を構成するサイクル数は、特に限定されない。

【0036】上記の方法で再帰的にコンセンサス・パスウェイを探索していく経路、階層構成は例えば図2のように様々なものが考えられる。各探索経路を左から1)serial search path、2)pyramid search path、3)combinatorial search pathと呼ぶことにする。コンセンサス・パスウェイとしての候補が複数ある場合はpyramid search path、もしくはcombinatorial search pathを用いることができる。serial search pathは、順次シリアルにコンセンサス・パスウェイを追加していく探索方法である。またpyramid search pathでは、階層ごとに候補として複数のコンセンサス・パスウェイを追加していく。Pyramid search pathに基づく探索方法は、図3のように模式化することができる。図3は、Pyramid search pathに基づく探索方法において、固定したネットワーク(コアネットワーク)に対して、考えうるネットワーク(デリバティブネットワーク)を生成する過程を示している。更にcombinatorial search pathは、各階層ごとに上位層のコンセンサス・パスウェイをまとめていく方法である。

【0037】続いて工程c)において、コンセンサス・パスウェイを含む新たなトポロジーが生成され、生成されたトポロジーに対してパラメータセットを与えることにより、二次的な候補ネットワークの集団が得られる。二次的な候補ネットワークの集団を生成するためには、前述の工程a)における候補ネットワークの生成手法をそのまま応用することができる。二次的な候補ネットワークの集団は、再び工程b)における候補ネットワークとして評価され、コンセンサス・パスウェイの抽出工程を繰り返される。2サイクル目以後に抽出されるコンセンサス・パスウェイは、それ以前に抽出されたコンセンサス・パスウェイに対して新規な構造を抽出しなければならない。すなわち、それ以前に抽出されたコンセンサス・パスウェイ以外の構造を抽出するか、またはそれ以前に抽出されたコンセンサス・パスウェイを含む、より構成要素の多い構造を抽出するかの、いずれかである。いずれの抽出方法を採用しても、最新のコンセンサス・パスウェイを含む全ての情報を総合して表現されるコンセンサス・パスウェイは等しい構造となる。

【0038】本発明では、候補ネットワークの生成と評価、そしてコンセンサス・パスウェイの抽出工程を繰り返すことによって、しだいにネットワーク空間が縮小す

る。すなわち、候補ネットワークのバリエーションが収束する。本発明のネットワーク推定方法は、ネットワーク空間が、所期の規模にまで縮小することができた段階で終了することができる。言いかえれば、与えられたデータを再現する可能性があるネットワークを、希望する大きさの集団に縮小できたときに、ネットワークの推定方法を終了することができる。

【0039】候補ネットワークのバリエーションが最小となるのは、工程c)において生成される二次的な候補ネットワーク空間の規模が、1サイクル前の規模と変わらなくなつたときである。このとき、その候補ネットワークの生成に用いたコンセンサス・パスウェイに対して与えられたパラメータと適応度からなるネットワークのバリエーションは最小である。したがって、本発明のネットワークの推定方法を、候補ネットワークのバリエーションが収束するまで繰り返せば、データを再現する可能性のある全てのネットワークを選び出すことができる。しかしながら、本発明のネットワーク推定方法は、ネットワーク空間の収束に至らない場合であっても、目的とする規模にそれを縮小することができれば、任意の段階で終了して良い。

【0040】上述のネットワーク推定方法を実行する装置および、上述のネットワーク推定方法をコンピュータ・システムに実現させることができるプログラムを格納した記録媒体も本発明に含まれる。

#### 【0041】

【実施例】本発明のネットワーク推定方法の有効性を示すために、以下の実験を行った。この推定方法においては、人為的なトリプレットの集合(トリプレットブル)の生成と、その中から候補トリプレットを抽出する処理の繰り返しで構成される。生成と抽出からなる1回の処理サイクルを、サンプリングサイクルと呼ぶ。この抽出処理には、構造近傍にあるネットワークの集合から、遺伝子ネットワークを構成する上で中心となるべき不可欠なトポロジーを導出する手法である「Iterative Sampling Method」を適用した。この手法では、候補ネットワークの集合からいかに信頼性の高いトポロジーを抽出するかに焦点を絞った点が大きな特徴である。本発明に基づくネットワークの推定方法のアルゴリズムを以下に示す。

1. 可能なネットワークトポロジーの生成、および各トポロジーについてターゲットデータと合致するようにパラメータの最適化を行う。
2. 各ネットワークの適応度を評価する(TSS等を利用)。
3. T個の適応度の高いネットワーク、もしくはウェイトをかけた全ネットワークからコンセンサス・パスウェイを導出する。
4. ある基準に沿ってコンセンサス・パスウェイを固定する。

5. 既に固定されているコンセンサス・パスウェイを含むネットワークトポロジーを生成し、各トポロジーについてパラメータを最適化する。

6. ステップ2へ戻る。

このアルゴリズムを図5に示す。ここで、コンセンサス・パスウェイはもっともらしいネットワークの要素として固定された結合構造であり、コアネットワーク(Core Network)とみなされる。各サイクルで一定数の結合構造が固定され、隨時「Fth-order core network」として一サイクル前に固定されたコアネットワークに追加される。Fはサンプリングサイクルの反復数を示す。

【0042】また、このアルゴリズムは、擬似コードを用いると以下のように示される。

```

F=1;
G(F)={φ};
while ( F ≤ Fmax ){
  D=G(F)を含むネットワークトポロジーの生成;
  D中の各ネットワークの適度度の評価(TSS等);
  D中のconsensus topologyの評価;
  if ( p が頻発 ){
    G(F+1)=G(F) ∪ {p};
  }
  F++;
}
G(F)を出力;

```

【0043】ここで、Dは候補ネットワークの集合、G(F)は階層の深さFでのコアネットワーク、pは候補グループの中でfunction size S(p)で定義された数をコア

10

20

ネットワークに追加されるコンセンサス・パスウェイをそれぞれ表している。このアルゴリズムの中でF=1のfirst-orderコアネットワークを生成する場合にはG(F)={φ}だから全くランダムにネットワークの候補グループが生成されDに格納される。ここから統計的な評価を行いpを抽出してfirst-orderコアネットワークとする。ここまでを一サイクルとし、さらにF=2のsecond-orderのネットワークを生成する場合には、first-orderコアネットワークを包含したデリバティブネットワークを候補グループとしてDに格納して評価を行い、pを上位のネットワーク、この場合はfirst-orderコアネットワークに追加してこれをsecond-orderコアネットワークとする。同様の作業をFがFmaxに達するまで繰り返す。つまり、最上位層のみをコアネットワークとするではなく、ある階層より下位のネットワークを探索する場合、その階層のネットワークをデリバティブネットワークと考え、さらにそこから機能を付加して下位層のコアネットワークを生成していく。

【0044】上記のG(F)のサイズを5000として、トリプレット抽出処理の実験を行った。TSSの分布状況を図6に示す。この中でエリート数E=300でpの抽出を行い、F=2までサイクルを回し、各サイクルともS(p)を2とした。一サイクル目の $\Gamma$ (activation) (式3)、 $\Gamma$ (inhibition) (式4)を以下に示す。下線を引いた要素が各行列の最大値である。実験で得られた階層構造を図7に示す。

【0045】

【数3】 (式3)

$$\Gamma(\text{activation}) = \begin{pmatrix} 20606.2 & \underline{62648.4} & 9184.12 & 16817.8 \\ 10298.4 & 3726.79 & 5923.83 & 8764.97 \\ 14187.5 & 19064.7 & 4204.16 & 8081.3 \\ 3241.13 & 13953.1 & 1082.62 & 5183.11 \end{pmatrix}$$

【0046】

$$\Gamma(\text{inhibition}) = \begin{pmatrix} 132.628 & 141.171 & 1082.3 & 4985.05 \\ 5448.63 & \underline{554.033} & 1878.69 & 1177.37 \\ 862.98 & 1622.94 & 1193.72 & 2394.87 \\ 898.286 & 11386 & 1519.07 & 0.279233 \end{pmatrix}$$

40

【0047】抽出された候補ネットワークの中には、図4(a)で示したターゲットプロファイルの一部である、AのA自身に対する活性、およびAのBに対する活性を示すネットワークを含む候補トポロジーが含まれていた。このことから、本発明により、効率的にターゲットプロファイルを生じるネットワーク構成を推定することが可能であることは明らかである。

【0048】具体的な遺伝子のネットワークを例に、上述の処理および候補トリプレットを抽出する処理を具体

的に説明する。ここでは、図4(b)に示したような時系列データである発現プロファイルを実現する遺伝子のネットワークを例とする。まず、トポロジーの結合関係を示す結合行列をもとに染色体にコーディングを行った。ここで言う染色体とは、遺伝的アルゴリズム用語としての染色体である。すなわち、遺伝子である結合荷重としきい値をコード化したものが、ここで言う染色体に相当する。染色体上にコード化されたパラメータに基づいてネットワークが構成される。この関係を生物学的用

50

語としての染色体に置きかえれば、染色体上にコード化されたパラメータは遺伝子に相当し、このパラメータに基づいて生成されるネットワークが発現型に相当することになる。ここでは、図4 (a) に示した発現物質が4つであるので、ノードが4つあるネットワークのトポロジーを考える。それは、例えば、以下のような結合行列Cで表される。

【0049】

【数5】

$$C = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \end{pmatrix}$$

【0050】このトポロジーの結合は5ヶ所あり、各々に対する重み付けをする必要があるので、4つの遺伝子型 (genotype) を持つ染色体を生成させた。さらに各ノードのしきい値も最適化する必要があるので、結局、全体として5 (結合数) + 4 (ノードのしきい値数) で9の遺伝子型を持つ染色体を生成させることになる。この場合の各染色体の荷重および、しきい値の取り得る範囲 (実数値) は、表1のように示される。

【0051】

【表1】

#### 結合荷重

A→A : [0~1]

A→B : [0~1]

B→C : [0~1]

C→D : [0~1]

D→B : [-1~0]

#### 各ノードのしきい値

A : [0~1]

B : [0~1]

C : [0~1]

D : [0~1]

以上のようにして、図4 (b) に示す発現プロファイルを実現可能な複数のトポロジーを生成して、トポロジー格納手段 (図5における第1オーダー、最も上のトポロジー格納手段) に格納した。

【0052】格納した各トポロジーに対して、GAを用いてパラメータ (荷重およびしきい値) の最適化を行った。まず、トポロジーに対して決められた範囲のパラメータをN組ランダムに発生させ、染色体の初期集団とし

た。表2に上述の結合行列で定められたトポロジーに対する初期集団を示す。表中の各個体は、5つの結合荷重および4つのしきい値を遺伝子とする染色体で構成されている。

【0053】

【表2】

結合荷重 個体1 個体2 個体3 個体N

10	A→A : 0. 2	0. 3	0	... 0. 3
	A→B : 0. 4	0. 7	0. 2	... 0. 2
	B→C : 0. 2	0. 2	0. 5	... 0. 6
	C→D : 0. 2	0. 4	0. 7	... 0. 2
	D→B : -0. 1	-0. 3	-0. 6	... -0. 5

しきい値

A :	0. 6	0. 6	0. 3	... 0. 2
B :	0. 2	0. 2	0. 3	... 0. 4
C :	0. 3	0. 4	0. 3	... 0. 6
D :	0. 2	0. 5	0. 3	... 0. 2

このように、トポロジーに対して、リストの形で染色体が定まる。この初期集団に対してGAを適用することにより、最適化を行った。この最適化の指標として、各パラメータの適応度を用いた。GAのパラメータは、population size=500、generation=300、crossover rate=0.8、mutation rate=0.01で行い、エリート保存戦略、トーナメント選択戦略を用いた。なお今回の実験では非結合の評価は行わないものとした。

【0054】初期集団の適応度は、染色体のパラメータをもとに、各個体で発現プロファイルを生成し、ターゲット・プロファイルとの間で、最小自乗誤差TSS (Total Sum Square error) の処理を行うことにより求めた。以下に最小自乗誤差TSS処理について述べる。

【0055】図4 (b) に示すように、ターゲット・プロファイルが濃度の時系列データとして与えられている。このため、各発現物質A、B、C、およびDそれぞれのある時間tにおける濃度は $L_{TA}(t)$ 、 $L_{TB}(t)$ 、 $L_{TC}(t)$ 、 $L_{TD}(t)$ と表され、生成した染色体の各パラメータをもとに、上述の式(1)により計算される各発現物質の濃度は、 $L_{EA}(t)$ 、 $L_{EB}(t)$ 、 $L_{EC}(t)$ 、 $L_{ED}(t)$ と表すことができる。これらの間の最小自乗誤差TSSは次のように表すことができる。このようにして求めた最小自乗誤差TSSを各発生したネットワーク (染色体) の適応度とした。

【0056】

【数6】

21

22

$$TSS = \sum_{X}^{A, B, C} \sum_{t=0}^T (L_{TX}(t) - L_{EX}(t))^2$$

【0057】さて、このように求められる適応度を利用することにより、GAを用いてトポロジー毎にパラメータ（結合荷重としきい値）の最適化を行うことができる。すなわち、各トポロジー毎に適応度に基づいて順位付けし、適応度の高いトポロジーを選択してトポロジー格納手段（第1オーダー、パラメータ最適化後のトポロジー格納手段）に格納することにより、最適化されたトリプレットを集めるトリプレットプールを構築することができる。構築されたトリプレットプールを対象として、先の計算式に基づいて $\Gamma$ の比較を行い、候補ネットワークを推定することができる。具体的には、まず $\Gamma_{(a, activation)}$ と $\Gamma_{(i, inhibition)}$ について、1サイクルで固定するコンセンサス・パスウェイの数量だけ、高い値を持つ上位の要素を抽出する。例えばそれぞれ1サイクルで1つを固定するならば、 $\Gamma_{(a)}$ と $\Gamma_{(i)}$ の中でそれぞれ最高値の要素を選び、それらをコンセンサス・パスウェイとする。2サイクル目以降は、順次その固定する数を増やしていく、これを必要なサイクル数だけ繰り返す。

【0058】（他の適用）上述の説明および実施例では、遺伝子ネットワークの例で説明した。この他にも例えば遺伝子・代謝回路ネットワークにも適用することができる。図8は、遺伝子・代謝回路のネットワークについて示している。図8（a）に示すように、遺伝子と、代謝における酵素、タンパク質をノードとして、活性、抑制、媒介等の関係をリンクで表す。そしてこれを用いたネットワーク例を図8（b）に示す。図8（b）に示したようなトポロジーに対して、上述で説明した処理を適用する。

【0059】上述においては、対象ネットワークとして、遺伝子ネットワーク、代謝回路ネットワークを示したが、そのほかにも神経回路ネットワーク等に対しても適用することができる。この場合のデータ（ターゲット・プロファイル）としては、ニューロン活動電位等を用いる必要がある。

【0060】このように、本発明の手法は色々なものに対して用いることができる。適用できるターゲット・プロファイルの色々な例を図9に示す。図9（a）は、ターゲット・プロファイルが0次元空間データである場合を示している。このデータの例としては時間や周波数に対して変化するようなデータである。例えば、濃度、活性度、電位等がある。図9（b）は、1次元空間データを示している。データは、空間として $x$ に、および、時間または空間に依存する量である。この量としては、濃度、活性度、電位等がある。図9（c）および（d）として示したターゲット・プロファイルは、2次元空間データ

一タおよび3次元空間データである。これは、2次元 $(x, y)$ および3次元 $(x, y, z)$ の空間、および、時間または周波数に依存する量のデータである。この量としては、濃度、活性度、電位等がある。このように色々なデータ形式で、ターゲット・プロファイルを与えることができる。

【0061】このように、ターゲット・プロファイルが存在し、そのターゲット・プロファイルを生じるネットワーク構成を推定するような場合に、本発明は有効である。本発明は、スタンド・アローンのコンピュータ・システムばかりでなく、複数のコンピュータから構成される例えばクライアント・サーバ・システム等に適用してもよい。本発明に関するプログラムを格納した記憶媒体から、プログラムをシステムで読み出して実行することにより、本発明の構成を実現することができる。この記憶媒体には、DVD、MD、MO、フロッピー・ディスク、CD-ROM、磁気テープ、ROMカセット等がある。

#### 【0062】

【発明の効果】上記の説明のように、本発明は、データとしてターゲット・プロファイルが存在し、そのターゲット・プロファイルを生じるネットワーク構成を推定するような場合に、極めて有効である。特に、遺伝子ネットワークや代謝ネットワークのように、分子生物学におけるモデルの推定に対して有効に適用することができる。

【0063】本発明は、少なくとも1つのデータプロファイルに基づくネットワークの推定を可能とする。したがって、データプロファイルを得るための実験工程を最小限にことができる。あるいは、mutantに基づくターゲット・プロファイルを組み合わせることによって、候補ネットワークの絞込みをより容易に行うことができる。また、本発明においては、必要に応じて複数の候補ネットワークの評価を容易に行うことができる。したがって、常に同定選択範囲を広くとらえることができ、結果としてより多様な可能性の解析が効率的に行われる。

【0064】本発明は、DNAチップを用いて得られた遺伝子発現プロファイルをターゲットプロファイルとするネットワークの推定方法に特に有用である。DNAチップには、ある一群の遺伝子に対するプローブ（プローブセット）が高密度に固定されている。同じプローブセットを固定したDNAチップでは、その他の条件を等しくすることによって再現性の高い解析結果を得ることができる。つまり、本発明におけるネットワークにおいて、共通のノードを持つネットワークに基づく発現プロ

ファイルが得られることになる。DNAチップを用いた発現プロファイルの解析は、異なる実験の間の解析結果を比較できるように、標準化のための基準も設定されている。たとえば、結果を大きく左右するハイブリダイゼーションや洗浄条件などは予め定められた共通の基準に適合させておくことにより、異なる実験の結果を容易に比較することができる。

【0065】さて、このようにして収集されたDNAチップによる解析結果は、まず第1の細胞の解析結果に基づいてネットワークの推定が行われる。更に、第2の細胞から得られた発現プロファイルに基づいてネットワークの推定が行われる。このとき用いたDNAチップが同じプローブセットを備えていれば、両者の発現プロファイルを統合して本発明によるネットワークの推定方法を行うことができる。このように、あるDNAチップを用いて得られた複数の発現プロファイルを蓄積し、蓄積された発現プロファイルに基づいて本発明のネットワークの推定方法を実施することにより、候補ネットワークの絞り込みを、より効率的に実施することができる。このような解析方法の最大のメリットは、さまざまな研究施設で行われた実験結果を一箇所に集積することによって、DNAチップを用いた解析結果を容易に統合できることにある。更に、同じノードで構成される多様な発現プロファイルが蓄積されることによって、さまざまな遺伝子の発現プロファイルを、コンピューターを用いて解析することができる。その結果、ある遺伝子を阻害した

10

場合の発現プロファイルをコンピュータ内で予測し、その予測を実験的に確認するために、同じDNAチップを用いた実験計画を構築することもできる。以上述べたように、本発明のネットワーク推定方法は、DNAチップを用いた遺伝子の発現プロファイルの解析において、特に有用である。

#### 【図面の簡単な説明】

【図1】分子生物学におけるコンピューター利用を説明する図である。

【図2】コンセンサス・パスウェイを探索する経路を示す図である。(1)はserial searchpath、(2)はpyramidal search path、(3)はcombinatorial search pathを表す。

【図3】コンセンサス・パスウェイを探索する経路であるpyramidal search pathの概念を示す図である。

【図4】ネットワークによって生成されたターゲットプロファイルを示す図である。(a)の数字は結合荷重値を表す。(b)横軸は時間を、縦軸は各遺伝子産物の濃度を表す。

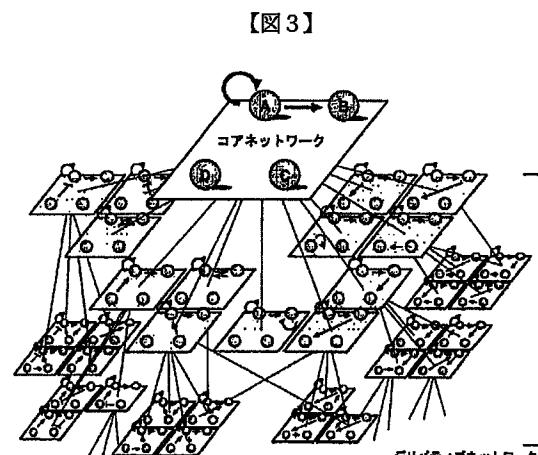
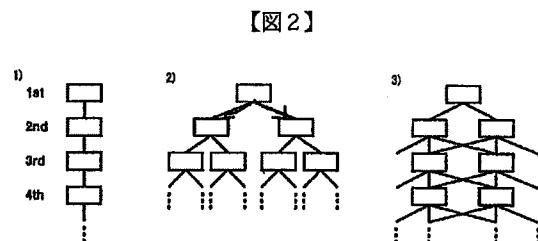
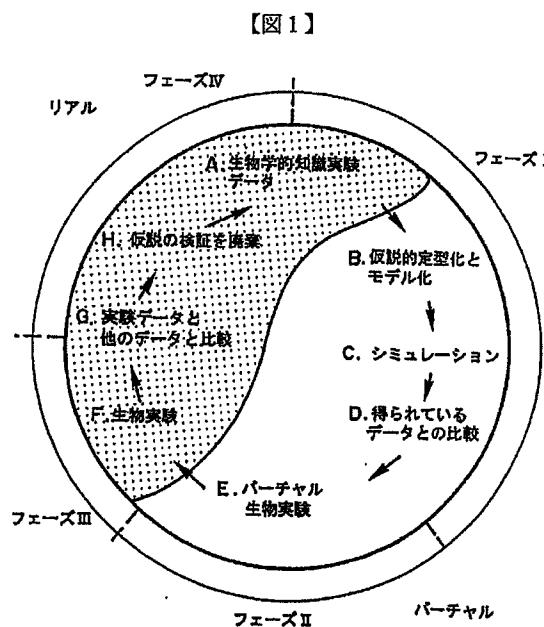
【図5】本発明によるネットワーク推定方法の例であるIterative Sampling methodの概念を示す図である。

【図6】最小自乗誤差TSSの分布状況を示す図である。

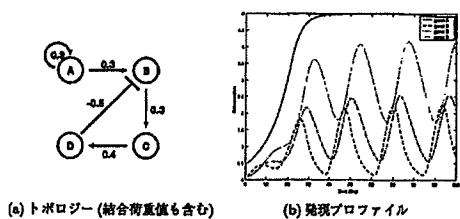
【図7】実験で得られた階層構造を示す図である。

【図8】遺伝子・代謝ネットワークの例を示す図である。

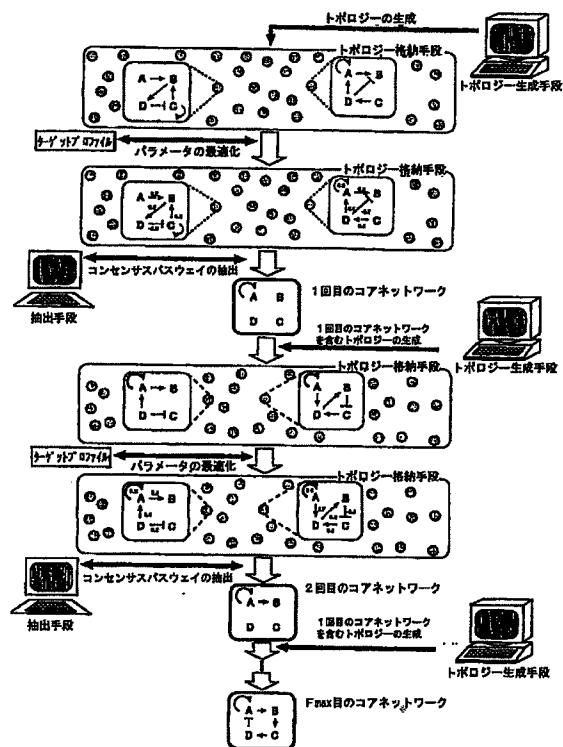
【図9】ターゲット・プロファイルの例を示す図である。



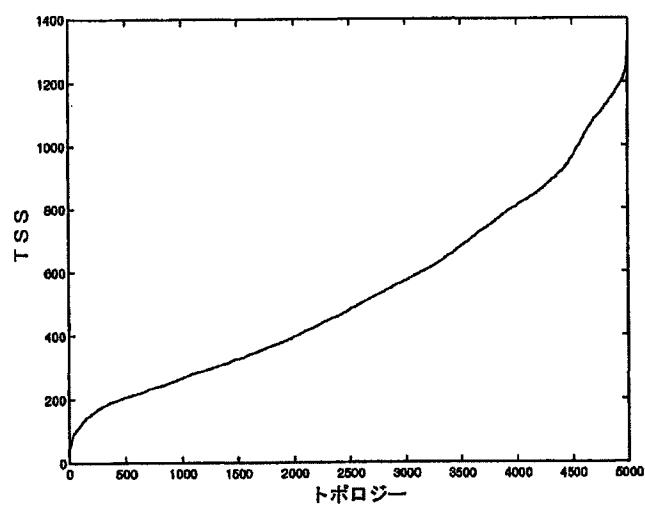
【图4】



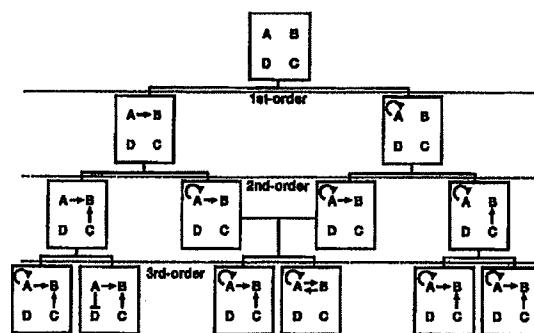
[图 5]



【图6】

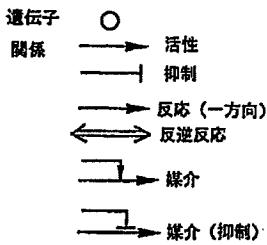


【図7】



【図8】

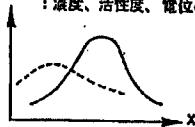
(a) 遺伝子代謝ネットワークの表示  
酵素, タンパク質 □



【図9】

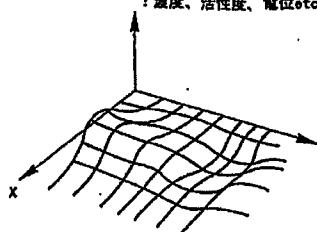
(a) 0次元空間データ

x : 時間もしくは周波数  
c : 濃度、活性度、電位etc



(b) 1次元空間データ

x : 空間  
c : 時間もしくは周波数  
c : 濃度、活性度、電位etc



(c) 2次元空間データ

x, y : 空間  
t : 時間もしくは周波数  
c : 濃度、活性度、電位etc

(d) 3次元空間データ

x, y, z : 空間  
t : 時間もしくは周波数  
c : 濃度、活性度、電位etc

(b) ネットワーク例

